SPAR: small RNA-seq portal for analysis of sequencing experiments

Pavel P. Kuksa[1,2,†], Alexandre Amlie-Wolf[1,2,3], Živadin Katanić[1,2], Otto Valladares[1,2], Li-San Wang[1,2,3,4] and Yuk Yee Leung[1,2,*,†]

[1]Penn Neurodegeneration Genomics Center, University of Pennsylvania, Philadelphia, PA 19104, USA, [2]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, [3]Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA and [4]Institute on Aging, University of Pennsylvania, Philadelphia, PA 19104, USA

ABSTRACT

The introduction of new high-throughput small RNA sequencing protocols that generate large-scale genomics datasets along with increasing evidence of the significant regulatory roles of small non-coding RNAs (sncRNAs) have highlighted the urgent need for tools to analyze and interpret large amounts of small RNA sequencing data. However, it remains challenging to systematically and comprehensively discover and characterize sncRNA genes and specifically-processed sncRNA products from these datasets. To fill this gap, we present *Small RNA-seq Portal for Analysis of sequencing expeRiments* (SPAR), a user-friendly web server for interactive processing, analysis, annotation and visualization of small RNA sequencing data. SPAR supports sequencing data generated from various experimental protocols, including smRNA-seq, short total RNA sequencing, microRNA-seq, and single-cell small RNA-seq. Additionally, SPAR includes publicly available reference sncRNA datasets from our DASHR database and from ENCODE across 185 human tissues and cell types to produce highly informative small RNA annotations across all major small RNA types and other features such as co-localization with various genomic features, precursor transcript cleavage patterns, and conservation. SPAR allows the user to compare the input experiment against reference ENCODE/DASHR datasets. SPAR currently supports analyses of human (hg19, hg38) and mouse (mm10) sequencing data. SPAR is freely available at https://www.lisanwanglab.org/SPAR.

INTRODUCTION

Increasing evidence of the significant regulatory roles of small non-coding RNAs (sncRNAs) in different diseases have been reported recently (1–3). Together with the introduction of new high-throughput small RNA sequencing protocols (4–6) that generate large-scale genomics datasets, there is a pressing need for tools capable of analyzing and interpreting large amounts of small RNA sequencing data, as well as discovering and characterizing all small RNA genes and specifically-processed small RNA products. Most existing tools (e.g., ShortStack (7), shortran (8), iMir (9), ncPRO-seq (10), FlaiMapper (11) or sRNAnalyzer (12)) focus on a specific sncRNA class such as miRNAs or only use existing annotations to quantify sncRNA expression. Thus, most of these tools are not designed for characterizing the full range of sncRNAs, which include many different classes, or to discover novel sncRNA loci contained in high-throughput sequencing data. On the other hand, previous tools identify transcriptionally active regions or read blocks and profiles (bgrSegmenter (13), blockbuster (14), SeRPeNT (15)) and do not focus on identifying peaks with the evidence of specific processing patterns, i.e. mature RNA products (e.g. miR-3p with low 5p end read entropy). Moreover, most existing tools for sncRNA-seq analyses such as FlaiMapper (11), SeRPeNT (15) support only BAM as the only input files, while ShortStack requires FASTA/FASTQ/BAM files and nc-PRO-seq accepts FASTQ/CSFASTA/BAM. Most importantly, no currently available online webserver can perform interactive analysis and visualization of custom small RNA sequencing datasets and compare them to publicly available reference small RNA datasets for downstream interpretation.

To address these gaps, we present the *Small RNA-seq Portal for Analysis of sequencing expeRiments* (*SPAR*) web server (https://www.lisanwanglab.org/SPAR). SPAR allows users to interactively analyze and visualize sncRNA sequencing data and individual small RNA loci from a

---

*To whom correspondence should be addressed. Tel: +1 215 573 3729; Fax: +1 215 573 3111; Email: yyee@pennmedicine.upenn.edu
†The first and last authors should be regarded as Joint First Authors.

variety of experimental protocols including smRNA-seq (16), microRNA-seq, short total RNA sequencing (4,5), and single-cell small RNA-seq (6). SPAR currently supports analyses of human (hg19, hg38) and mouse (mm10) genomes. SPAR also includes publicly available reference sncRNA datasets from DASHR (16) and ENCODE (4) across 185 tissues and cell types to produce highly informative small RNA annotations across all major small RNA types. Thus, SPAR is the first webserver that enables users to analyze, visualize and seamlessly link their own sequencing data with reference small RNA datasets from ENCODE and DASHR. The user provides URLs to web-accessible indexed BAM (read alignment data) or BigWig (genome-wide read coverage) small RNA-seq files or selects one of the public datasets from the ENCODE or DASHR databases. SPAR performs unsupervised segmentation to identify sncRNA loci displaying features of specific processing in the provided sequencing data. SPAR then characterizes these discovered loci using comprehensive RNA annotations across nine sncRNA classes and other features such as co-localization with other genomic features, precursor transcript cleavage patterns, and conservation. SPAR provides visualization and summarization of genome-wide and specific sncRNA results in the following formats: interactive sncRNA loci table (Javascript), genome browser, plots in PNG, downloadable BED, BigWig, BigBed tracks, and XLS spreadsheets. SPAR compares the user's sncRNA results with any of the public experimental data from DASHR and ENDCODE by matching detected loci against sets of loci identified in the reference datasets for each tissue and cell type (XLS format).

SPAR WORKFLOW

The SPAR web server provides a unique set of features for interactive analysis and visualization of small RNA sequencing datasets. The SPAR workflow is comprised of four components (Figure 1).

The 'Data Processing' component identifies sncRNA loci displaying features of specific processing from user-provided data (URLs to web-accessible indexed BAM or BigWig small RNA-seq files) or any reference datasets from the ENCODE (4) or DASHR (16) databases using unsupervised segmentation (16) (see also Supplementary Methods and Supplementary Figure S1 for accuracy evaluation). The SPAR pipeline is highly efficient: analysis and annotation of 100M reads takes <60 s and <2GB RAM (1.2M reads/s).

The 'Characterization of sncRNAs' component annotates the identified sncRNA loci across all major small RNA types (microRNA (miRNA), Piwi-interacting RNA (piRNA), transfer RNA (tRNA), tRNA fragments (tRFs), small nuclear (snRNA), small nucleolar (snoRNA), small cytoplasmic (scRNA), and ribosomal RNA (rRNA)) (16) or as novel, previously un-annotated RNAs (see Supplementary Methods). Each sncRNA locus is also characterized with a comprehensive set of features including conservation, co-localization with various genomic features (promoters, UTRs, introns, exon, lncRNAs, repeats) and features that are representative of RNA processing patterns (consistency of 5′/3′ processing, length, etc.) (see Supplementary Methods for details).
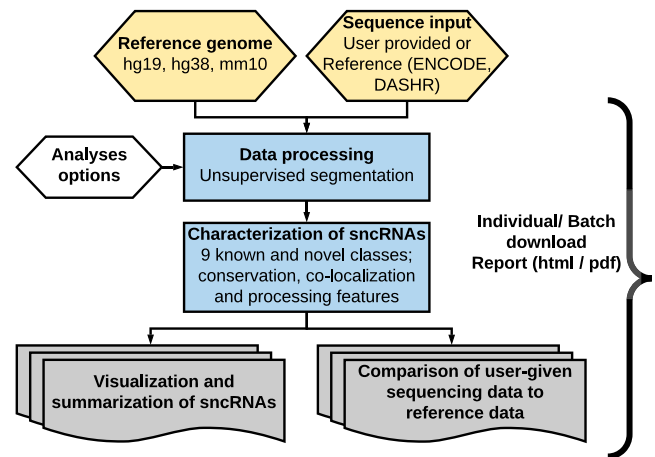

Figure 1. The SPAR workflow. Given a reference genome and input small RNA-seq dataset (custom or reference data), SPAR processes the small RNA-seq dataset and identifies sncRNA loci using unsupervised segmentation. sncRNA loci are grouped into the major small RNA classes or the novel unannotated category (total of 10 classes) and annotated with various genomic features. The SPAR results are summarized in interactive tables at the per locus and genome-wide level and are compared with reference ENCODE and DASHR tissues and cell lines. All results are available for download in a variety of formats.

The 'Visualization and summarization of sncRNAs' component provides summary tables, interactive sncRNA loci tables (Javascript), and genome-wide plots of sncRNAs (PNG and genome browser view) to characterize the quality of the data and results. All tables, figures, and tracks in the SPAR output are available for download individually or as a report (PDF).

Finally, the 'Comparison of user-given sequencing data to reference data' component allows users to compare their data with the collection of curated and processed public small RNA sequencing datasets from DASHR and ENCODE to provide novel biological insights. SPAR provides a table comparing detected loci in the input experiment against the sets of loci identified in the reference datasets across 185 tissues and cell types (expression values (reads per million, RPM) (17), XLS format).

SPAR WEB SERVER

Input

Sequence and reference genome. The user provides a small RNA sequencing dataset as input. Under 'Analyze your own data' tab, the user can provide a small RNA dataset as custom input in an indexed BAM (read alignment data) or BigWig (genome-wide read coverage file) formats (Figure 2). The user can directly upload the data file or specify web-accessible URLs (Figure 2). Under 'Analyze public datasets' tab, the user can select from a drop-down menu any of the 185 reference datasets from the ENCODE and DASHR databases for analyses (Figure 2). SPAR currently supports reference genomes for human (hg19, hg38) and mouse (mm10) which can be selected in the simple drop-down menu.

Figure 2. SPAR inputs. SPAR can be used to interactively process and analyze small RNA sequencing datasets. The user can select the input dataset from one of the publicly available DASHR/ENCDODE datasets ('Analyze public datasets', left box) or provide their own custom data via web-accessible URL or direct upload ('Analyze your own data', right box). SPAR parameters can be specified under 'Additional analysis options'. Genomic regions of interest can be provided by the user under 'Specify regions of interest' section.

Analysis options. 'Additional analysis options' section (Figure 2) allows users to specify and adjust parameters used by the SPAR segmentation algorithm to detect sncRNA loci (e.g. minimum and maximum lengths for detected loci, minimum read coverage, etc). SPAR parameters are explained in more details the Supplementary Methods. 'Specify regions of interest' section (Figure 2) allows users to upload custom intervals specifying genomic regions of interest in the BED/BED6/BED12/GFF or GTF file format, to restrict the SPAR analyses on these regions only.
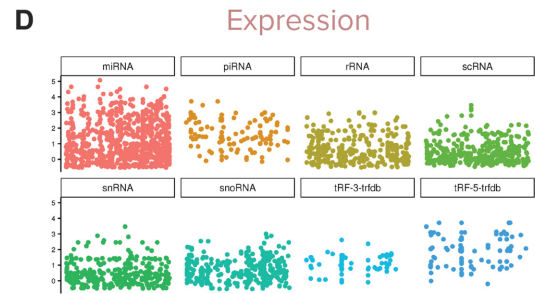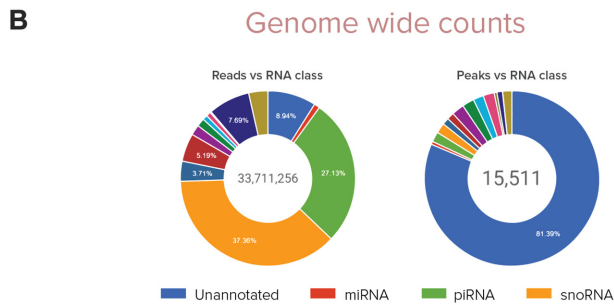
Output

Report page. After the user submits the input data, a new tab (report page) is automatically displayed, showing progress indicators for each major step in the SPAR workflow. Upon completion, the results are displayed on a report page organized into several sections:

1) Report section: results can be downloaded as a single ZIP archive and a URL is provided to access the SPAR results on the server for 2 weeks (Figure 3A).
2) The 'Run summary' section summarizes the overall composition of the small RNA-seq data, including the total number of detected sncRNA loci, the sncRNA

classes of the detected loci, the proportions of annotated and novel loci, their average length, expression levels, and other characteristics of the loci (Figure 3B).
3) The 'Interactive Peak Browser' is an interactive table that allows user to view, sort, and filter small RNA loci with >50 features including expression (RAW, RPM, percentile), conservation, co-localization with other genomic features, RNA class, processing specificity, etc. User can view individual peak information in the UCSC genome browser by clicking the 'peakID' they are interested in (Figure 3C).
4) The 'View in Genome browser' section displays the integrated JBrowse genome browser including annotation, peak and raw signal tracks. Users can also view these tracks by clicking the UCSC Genome browser link.
5) The 'Plots' section contains output figures in PNG format. Output plots are sub-divided into four categories: 'Segmentation' summarizes the characteristics of the loci before annotation; 'Length/expression' summarizes their characteristics after annotation, including their length distributions and expression levels (Figure 3D); 'Processing' summarizes the specificity of post-transcriptional processing of sncRNA transcripts (Figure 3E); and 'Genome-wide' summarizes sncRNA char-

A Link to results: https://www.lisanwanglab.org/SPAR/SPAR_out/fe73594 (will be kept for 2 weeks)
Link to results: Download results (ZIP) (will be kept for 2 weeks)
Link to report (PDF): Download report (PDF)

B Genome wide counts

Reads vs RNA class

Peaks vs RNA class

33,711,256

15,511

Unannotated | miRNA | piRNA | snoRNA

D Expression

miRNA | piRNA | rRNA | scRNA

snRNA | snoRNA | tRF-3-trfdb | tRF-5-trfdb

C Interactive sncRNA loci table

Loci: 1-10 / 15511   Page 1 of 1552   Results per page 10   ?   Columns▾

peakID	Same 5' end read %	Expression (RPM)	Annotation	Annotated class	Conservation	mRNA?	lncRNA?	repeat?
				Clear		Clear	Clear	Clear
P376	0.394643	1.3695	chr1:108113071-108113633:-:LSU-rRNA_Hsa	rRNA	0.0142	Intergenic	Non-lncRNA	Non-repeat
P375	0.920336	5.7629	chr1:108113071-108113633:-:LSU-rRNA_Hsa	rRNA	0.0043	Intergenic	Non-lncRNA	Non-repeat
P374	0.163942	3.4948	chr1:108113071-108113633:-:LSU-rRNA_Hsa	rRNA	0.0086	Intergenic	Non-lncRNA	Non-repeat
P315	0.448352	18.4211	chr1:109643154-109643236:+:HBII-382	snoRNA	0.9413	Intergenic	lncRNA Exon	Non-repeat

E Specific processing

miRNA | sc-sn-sno-RNA | tRFs

-1.0   -0.5   0.0   0.5   1.0
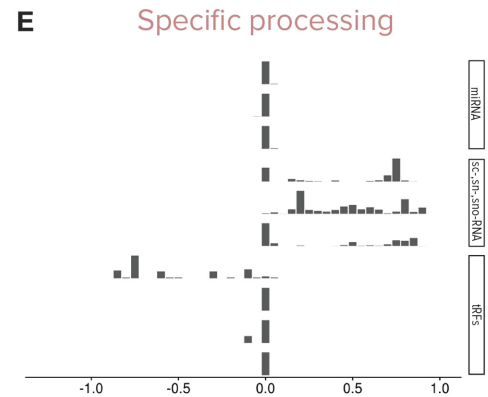
F Comparison of user sequencing data to reference data

	User data					DASHR data				
peak Chr	peak ChrStart	peak ChrEnd	peakID	peak Expression	peak Strand	adipose1	bladder1	brainog1	brainpfc1	brainpfc2
chr1	566410	566441	P1026	47183	-	727.5	NA	271.9	142809	1156.3
chr1	1102503	1102524	P15	17	+	NA	1524.5	NA	11	NA
chr1	1102539	1102562	P16	281	+	148.8	15433.9	NA	223	156.7
chr1	1103257	1103278	P17	21	+	NA	319.3	NA	NA	15
chr1	1103295	1103318	P18	4243	+	NA	9618.3	NA	214	138.9

Figure 3. SPAR outputs: (**A**) results can be downloaded easily as a single file (ZIP archive) or report (PDF). Individual files are available for download under 'Download results' section on the report page; (**B**) summary of the overall composition of the small RNA-seq data; (**C**) SPAR allows user to view, filter and download selected loci using the interactive sncRNA table; (**D**) SPAR provides overview of expression patterns across all loci in the small RNA sequencing data; (**E**) SPAR displays the variety of specific processing patterns across all loci in the small RNA sequencing data; (**F**) SPAR provides direct comparison of user data against ENCODE/DASHR reference tissues (RPM expression values).

acteristics across the whole genome, including a high-level view of co-localization of the loci with genomic features.

6) The 'Download results' section allows users to download results for specific RNA classes or across all sncRNA loci in BED, BigWig and BigBed formats, and XLS spreadsheets (see 'Downloadable SPAR outputs' section).

7) The 'Run log' section stores the information on how the SPAR job was performed. Timestamps, input files, parameters used and job ID are displayed.

Categorization of sncRNA loci in the SPAR output. SPAR characterizes discovered sncRNA loci using comprehensive sncRNA gene and mature product annotations in DASHR (16) including miRNAs (miRNAprimary, miR-3p, miR-5p), piRNAs, snRNAs, snoRNAs, scRNAs, tRNAs, tRFs

and rRNAs. Additionally, tRFs from the tRFdb database (18), newly described snoRNA mature products from the snoRNAome database, a recent experimental study (19), and piRNA records from piRNAbank (20) are included. sncRNA loci not overlapping existing annotations are categorized as novel and un-annotated loci (e.g. Figure 3B).

Characteristics of sncRNA loci. For each sncRNA locus, SPAR generates features for (i) precursor transcript cleavage patterns (e.g. processing specificity as [1-normalized entropy of 5′ reads ends across the alternative 5′ end positions]): the higher the number, the more specifically processed as opposed to randomly cleaved the locus is; (ii) co-localization with non-small RNA genes and other genomic elements such as repeat elements, promoters, exons, introns, lncRNA exons, lncRNA introns: IDs and coordinates of each co-localized element are returned and (iii) evolutionary conservation scores derived from the UCSC phastCons conservation tracks for human and mouse (21,22). Details of these computations can be found in Supplementary Methods, DASHR (16) and in the 'About' page on the SPAR website. SPAR allows the user to filter and sort sncRNA loci using one or more of these features (Figure 3C).

Comparison of the input/user-given sequencing data to reference datasets. To facilitate comparison and downstream analyses of sncRNA sequencing results, SPAR provides head-to-head comparison of sncRNA loci in the input experiment with all publicly available experimental data from DASHR and ENCODE. SPAR outputs a set of comparison tables generated by matching the set of detected loci in the user experiment against sets of loci identified in each tissue and cell type in the reference datasets (Figure 3F) and reporting RPM expression values for comparison.

Downloadable SPAR outputs. All results files generated by SPAR are available for download under the 'Download results' section of the report page. Users can download results for specific RNA classes or across all sncRNA loci as XLS spreadsheets. These files will contain sncRNA loci annotations, peak information as well as various features and scnRNA characteristics including per-tissue expression, conservation, specificity of sncRNA processing, and co-localization. Spreadsheets containing comparison of the input/user-given sequencing data to reference datasets are also available. Other files available for download include plots in PNG format, genome-wide raw coverage, peak, annotation tracks in BED, BigWig and BigBed formats.

EXAMPLE USES

We describe three example use cases for the SPAR web server:

1. Annotate and visualize sncRNAs from small RNA sequencing experiment
2. Perform comparative expression analysis to identify disease or tissue-specific loci
3. Identify top sncRNAs expressed in a particular tissue

Corresponding figures and tutorials are provided at the SPAR web site under the 'Tutorial' section. In the following, we describe the use of SPAR for comparative expression analysis (Use case 2). Additionally, we provided a tutorial (under "Preparing files for SPAR" section in tutorials https://www.lisanwanglab.org/SPAR/tutorials) on how to prepare bigWig and BAM files for analysis with the SPAR webserver. We have also created a repository (http://bitbucket.org/wanglab-upenn/spar_prepare) with the scripts that can be used to process user's sequencing data and generate the necessary files for analyses with the SPAR webserver.

Using small RNA sequencing data from Alzheimer's disease (AD) brain tissues, the goal is to identify sncRNAs that are expressed only in the AD data and other reference brain tissues in DASHR but not in any blood related tissues, including plasma, serum or whole blood. Corresponding figures can be found in the Tutorial page on the SPAR website.

To perform this analysis, first, we clicked 'analyze your own data' tab and chose 'hg19' under 'Genome'. Under 'Type of Input', we selected 'Raw signal (BigWig)', and provided two URLs for BigWigs with raw read coverage from our small RNA sequencing experiment, one for each strand. Then, we uploaded a BED file containing 21 genes of interest that contain SNPs associated with AD from the IGAP project (23) in the 'Specify regions of interest' section. We then clicked 'Analyze your own BigWig' to start SPAR analyses.

The resulting outputs were displayed in the report page (new tab), including job IDs, summary figures, interactive peak browser and genome browser, plots, downloads and run logs. SPAR results overlapping the genes of interest in the provide BED file are available under the 'Custom annotation summary' section in XLS format. This XLS table contains only the sncRNA loci identified in our AD brain data that overlapped with the IGAP genes.

To search for potential sncRNA loci expressed solely in AD brain, we downloaded the 'DASHR_comparison_annot.xls' table under 'Comparison with DASHR'. The first 28 columns of this table contain SPAR features for annotated loci in the AD sequencing data and the remaining 54 columns display the expression (RPM) for the same loci across DASHR tissues. To achieve our goal, we removed any DASHR reference tissues except the brain- (brainog1_SRA012516, brainpfc1_GSE43335, brainpfc2_GSE48552, braintgm1_GSE46131) and blood- (plasma1_GSE52981, serum1_GSE53439, serum2_GSE34891 and wholeblood1_GSE46579) related samples. We then retained loci that were not expressed ('NA' values) in the blood samples and expressed in the reference brain samples. A total of 63 sncRNA loci including 41 miRNAs, 7 piRNAs, 5 sno/snRNAs, 7 tRFs and 3 tRNAs remained. Out of these, 33 miRNAs have been reported as an AD biomarker (24–26); others will serve as candidate biomarkers for further biological interpretation.

DESIGN AND IMPLEMENTATION

SPAR internal programs are implemented using BASH, C, AWK, JavaScript and R scripts. Apache web server, PHP and Javascript are used for the web interface. The R package ggplot2 is used for visualization. The entire system is de-

ployed on a server with 32GB of RAM and eight 2.6 GHz Xeon CPUs. The performance of such implementation has been shown to be able to deal with ∼100s of runs on a daily basis, based on the performance of our other tools deployed with the same configurations. SPAR has been tested with major modern browsers such as Google Chrome, Mozilla Firefox, Microsoft Edge, and Safari, and the internal programs were used to process and annotate small RNA sequencing datasets in our recent publications (16).

CONCLUSIONS

With the increasing evidence of the significant regulatory roles of sncRNAs and growing number of large-scale small RNA-seq datasets generated by a variety of experimental protocols (Illumina small RNA-seq, ENCODE short total RNA-seq, ENCODE microRNA-seq, single cell small RNA-seq) their systematic interpretation and comparison becomes more challenging but critically important. With SPAR, users can easily and quickly analyze, annotate, visualize and compare small RNA sequencing data. With a unique and user-friendly interface, SPAR is the first web-server that provides users the ability to streamline the interpretation of small RNA-seq results and compare with hundreds of publicly available datasets.

We plan to extend SPAR with the integration of other small RNA datasets (ENCODE microRNA-seq, single-cell RNA datasets) and to support analyses for other model organisms. Additional analysis modules including differential expression analysis, target gene analyses, miRNA isoform detection and quantification, and functional characterization of the novel loci (15,27) will be included, thus further aiding the interpretation of genome-scale sncRNA data.

DATA AVAILABILITY

SPAR webserver is available at https://www.lisanwanglab.org/SPAR. Additionally, SPAR can be run using command line interface and scripts freely available at https://bitbucket.org/account/user/wanglab-upenn/projects/SPAR.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

REFERENCES

1. Goodarzi,H., Nguyen,H.C.B., Zhang,S., Dill,B.D., Molina,H. and Tavazoie,S.F. (2016) Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell.*, **165**, 1416–1427.
2. Soares,A.R. and Santos,M. (2017) Discovery and function of transfer RNA-derived fragments and their role in disease. *Wiley Interdiscip. Rev. RNA.*, **8**, e1423.
3. Steinbusch,M.M.F., Fang,Y., Milner,P.I., Clegg,P.D., Young,D.A., Welting,T.J. and Peffers,M.J. (2017) Serum snoRNAs as biomarkers for joint ageing and post traumatic osteoarthritis. *Sci. Rep.*, **7**, 43558.
4. Sloan,C.A., Chan,E.T., Davidson,J.M., Malladi,V.S., Strattan,J.S., Hitz,B.C., Gabdank,I., Narayanan,A.K., Ho,M., Lee,B.T. *et al.* (2016) ENCODE data at the ENCODE portal. *Nucleic Acids Res.*, **44**, D726–D732.
5. Djebali,S., Davis,C.A., Merkel,A., Dobin,A., Lassmann,T., Mortazavi,A., Tanzer,A., Lagarde,J., Lin,W., Schlesinger,F. *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.
6. Faridani,O.R., Abdullayev,I., Hagemann-Jensen,M., Schell,J.P., Lanner,F. and Sandberg,R. (2016) Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.*, **34**, 1264–1266.
7. Axtell,M.J. (2013) ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA*, **19**, 740–751.
8. Gupta,V., Markmann,K., Pedersen,C.N.S., Stougaard,J. and Andersen,S.U. (2012) shortran: a pipeline for small RNA-seq data analysis. *Bioinformatics*, **28**, 2698–2700.
9. Giurato,G., De Filippo,M., Rinaldi,A., Hashim,A., Nassa,G., Ravo,M., Rizzo,F., Tarallo,R. and Weisz,A. (2013) iMir: an integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-Seq. *BMC Bioinformatics*, **14**, 362.
10. Chen,C.-J., Servant,N., Toedling,J., Sarazin,A., Marchais,A., Duvernois-Berthet,E., Cognat,V., Colot,V., Voinnet,O., Heard,E. *et al.* (2012) ncPRO-seq: a tool for annotation and profiling of ncRNAs in sRNA-seq data. *Bioinformatics*, **28**, 3147–3149.
11. Hoogstrate,Y., Jenster,G. and Martens-Uzunova,E.S. (2015) FlaiMapper: computational annotation of small ncRNA-derived fragments using RNA-seq high-throughput data. *Bioinformatics*, **31**, 665–673.
12. Wu,X., Kim,T.-K., Baxter,D., Scherler,K., Gordon,A., Fong,O., Etheridge,A., Galas,D.J. and Wang,K. (2017) sRNAnalyzer-a flexible and customizable small RNA sequencing data analysis pipeline. *Nucleic Acids Res.*, **45**, 12140–12151.
13. Habegger,L., Sboner,A., Gianoulis,T.A., Rozowsky,J., Agarwal,A., Snyder,M. and Gerstein,M. (2011) RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. *Bioinformatics*, **27**, 281–283.
14. Langenberger,D., Pundhir,S., Ekstrøm,C.T., Stadler,P.F., Hoffmann,S. and Gorodkin,J. (2012) deepBlockAlign: a tool for aligning RNA-seq profiles of read block patterns. *Bioinformatics*, **28**, 17–24.
15. Pagès,A., Dotu,I., Pallarès-Albanell,J., Martí,E., Guigó,R. and Eyras,E. (2018) The discovery potential of RNA processing profiles. *Nucleic Acids Res.*, **46**, e15.
16. Leung,Y.Y., Kuksa,P.P., Amlie-Wolf,A., Valladares,O., Ungar,L.H., Kannan,S., Gregory,B.D. and Wang,L.S. (2016) DASHR: database of Small human noncoding RNAs. *Nucleic Acids Res.*, **44**, D216–D222.
17. Anders,S., McCarthy,D.J., Chen,Y.S., Okoniewski,M., Smyth,G.K., Huber,W. and Robinson,M.D. (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.*, **8**, 1765–1786.
18. Kumar,P., Mudunuri,S.B., Anaya,J. and Dutta,A. (2015) tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res.*, **43**, D141–D145.
19. Jorjani,H., Kehr,S., Jedlinski,D.J., Gumienny,R., Hertel,J., Stadler,P.F., Zavolan,M. and Gruber,A.R. (2016) An updated human snoRNAome. *Nucleic Acids Res.*, **44**, 5068–5082.
20. Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, **36**, D173–D177.
21. Siepel,A., Bejerano,G., Pedersen,J.S., Hinrichs,A.S., Hou,M., Rosenbloom,K., Clawson,H., Spieth,J., Hillier,L.W., Richards,S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
22. Tyner,C., Barber,G.P., Casper,J., Clawson,H., Diekhans,M., Eisenhart,C., Fischer,C.M., Gibson,D., Gonzalez,J.N., Guruvadoo,L. *et al.* (2017) The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res.*, **45**, D626–D634.
23. Lambert,J.C., Ibrahim-Verbaas,C.A., Harold,D., Naj,A.C., Sims,R., Bellenguez,C., DeStafano,A.L., Bis,J.C., Beecham,G.W., Grenier-Boley,B. *et al.* (2013) Meta-analysis of 74,046 individuals

identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.*, **45**, 1452–1458.

24. Santa-Maria,I., Alaniz,M.E., Renwick,N., Cela,C., Fulga,T.A., Van Vactor,D., Tuschl,T., Clark,L.N., Shelanski,M.L., McCabe,B.D. *et al.* (2015) Dysregulation of microRNA-219 promotes neurodegeneration through post-transcriptional regulation of tau. *J. Clin. Invest.*, **125**, 681–686.

25. Lau,P., Bossers,K., Janky,R., Salta,E., Frigerio,C.S., Barbash,S., Rothman,R., Sierksma,A.S., Thathiah,A., Greenberg,D. *et al.* (2013) Alteration of the microRNA network during the progression of Alzheimer's disease. *EMBO Mol. Med.*, **5**, 1613–1634.

26. Fang,M., Wang,J., Zhang,X., Geng,Y., Hu,Z., Rudd,J.A., Ling,S., Chen,W. and Han,S. (2012) The miR-124 regulates the expression of BACE1/β-secretase correlated with cell death in Alzheimer's disease. *Toxicol. Lett.*, **209**, 94–105.

27. Leung,Y.Y., Ryvkin,P., Ungar,L.H., Gregory,B.D. and Wang,L.-S. (2013) CoRAL: predicting non-coding RNAs from small RNA-sequencing data. *Nucleic Acids Res.*, **41**, e137.